8.0 OUTLIER ANALYSIS

In this section, the outlier analysis is discussed. First is a discussion of the general approach to the analysis, followed by details on how the data were grouped, a description of the outlier analysis procedure used, and a discussion of how the outliers found were handled in the statistical analysis. Data from House 08 (the house for which no pre-sampling XRF measurements were taken), which were excluded from the full statistical analysis, were included in this outlier analysis.

8.1 APPROACH

Formal statistical outlier tests were performed on both the field data and the laboratory QC data. Data were placed into groups for comparable types of samples, and a maximum absolute studentized residual procedure was used to identify potential outliers. When a potential outlier was identified, that value was excluded from the group, and the outlier test was performed again. This procedure was repeated until no additional outliers were detected. After all potential outliers were identified, a list of these samples was sent to the laboratory for rechecking.

8.2 DATA GROUPS

Samples collected from inside the houses were grouped according to the predominant interior abatement method, sampling method (vacuum or wipe) and component (air duct, floor, window channel, field blank, trip blank, etc.). Soil samples and exterior entryway vacuum samples were grouped according to the predominant exterior abatement method. In addition, interior floor samples were split into two groups, those taken from carpeted floors and those taken from uncarpeted floors. Separate outlier analyses were then performed for each group on the natural logarithm of lead loading values, the natural logarithm

of lead concentration values, sample concentration values (field blanks only) and net weight values (trip blanks only).

Normally, foundation soil samples were collected from the soil along the foundation of each house. In one case, however, pavement along the foundation required the use of a vacuum cassette to collect two dust samples rather than the usual two soil samples. Additional outlier tests were performed (1) grouping these two samples with foundation soil samples, and (2) grouping these two samples with exterior entryway vacuum samples.

Laboratory QC data were grouped according to type of sample and sample medium. Outlier analyses were then performed on the natural logarithm of the appropriate measurement for each type of sample (spike recovery for spiked samples; amount of lead for method blanks, calibration blanks, and unspiked samples; percent recovery for interference check samples, calibration standards, calibration verification samples and blind reference material samples; and range of spike recovery for duplicate spiked samples).

8.3 THE OUTLIER TEST

The SAS procedure GLM (SAS PC, ver. 6.04) was used to compute the studentized residual for each data value by fitting a "constant" model (i.e., mean value plus error term) to the log-transformed data in each group. The absolute values of the studentized residuals were then compared to the upper .10/n quantile of a t distribution with n-2 degrees of freedom, where n was the number of data values in the group. If the maximum absolute studentized residual was greater than or equal to the .10/n quantile, the corresponding data value was flagged as a potential outlier. The outlier test was then repeated, excluding additional potential outliers, until no more outliers were detected. Table 8-1 lists the field sample outliers found as a

result of this test. Table 8-2 lists the laboratory QC sample outliers.

Table 8-1. CAP Study Outliers - Field Samples

Lead Loading Outliers

Instrument Batch	Sample Preparation Batch	Lab ID	Sample Medium	Study ID/ Sample ID	Location	Component	Lead Loading[a] (ug/ft^2)
E04292A	WIO	902924	Dust-Vacuum	28/01	Kitchen	Floor	< 0.34
E05072B	WIR	903347	Dust-Vacuum	96/02	Hall	Floor	2365.43
E05072B	WJB	903556	Dust-Vacuum	19/01	Living Room	Floor	1102.35
E05132A	WJC	903116	Dust-Vacuum	96/01	Hall	Floor	11641.25
E06022A	WJG	902546	Dust-Vacuum	45/07	Kitchen	Floor	1765.38
E07272A	WIZ	903392	Dust-Vacuum	19/02	Living Room	Floor	6745.20
E07272A	WIZ	903769	Dust-Vacuum	21/25	Laundry Room	Floor	7046.70
E08032A	WKF	905079	Dust-Wipe	21/26	Laundry Room	Floor	333.56
E08032A	WKG	905143	Dust-Wipe	57/27	Bathroom #2	Floor	< 2.72

Lead Concentration Outliers

Instrument Batch	Sample Preparation Batch	Lab ID	Sample Medium	Study ID/ Sample ID	Location	Component	Lead Concentration[a] (ug/g)
E04272A	WIL	902564	Dust-Vacuum	17/13	Front	Outside Entryway	8.84
E04292A	WIL	902761	Dust-Vacuum	94/12	Hall	Inside Entryway	21.67
E04292A	WIO	903673	Dust-Vacuum	46/05	Bathroom	Air Duct	4623.43
E05072B	WIR	902605	Dust-Vacuum	79/12	Kitchen	Inside Entryway	2723.16
E05072B	WIR	903347	Dust-Vacuum	96/02	Hall	Floor	1724.32
E05072B	WJD	902142	Dust-Vacuum	49/02	Kitchen	Floor	< 4.56
E05072B	WJD	903487	Dust-Vacuum	60/01	Bedroom #1	Floor	< 11.00
E05122B	WJE	902126	Dust-Vacuum	79/14	Back	Outside Entryway	16335.45
E05122B	WJF	902220	Dust-Vacuum	51/02	Bathroom	Floor	13567.76
E05132A	WJC	903116	Dust-Vacuum	96/01	Hall	Floor	6217.62
E05192A	WIQ	904271	Soil	81/17	Back	Foundation	3351.12
E05262A	WIT	904054	Soil	79/16	Back	Entryway	< 4.55
E06022A	WJG	902546	Dust-Vacuum	45/07	Kitchen	Floor	6398.60
E06042A	WJP	902380	Dust-Vacuum	68/10	Dining Room	Air Duct	5644.54
E06112A	WIW	904433	Soil	51/18	Back	Foundation	< 5.49[1]
E06122A	WJR	903291	Dust-Vacuum	72/11	Hall	Inside Entryway	9.65
E06152A	WJV	903089	Dust-Vacuum	68/12	Kitchen	Inside Entryway	1200.39
E06292A	WKB	902955	Dust-Vacuum	80/11	Living Room	Inside Entryway	5332.00
E06292A	WKB	903020	Dust-Vacuum	03/04	Bathroom	Window Stool	48271.93
E06292A	WKB	903163	Dust-Vacuum	31/07	Bathroom #2	Floor	1.71
E07212A	WJG	902953	Dust-Vacuum	51/01	Bathroom	Floor	12186.30
E07212A	WJR	902169	Dust-Vacuum	19/12	Kitchen	Inside Entryway	2293.62
E08242A	WJA	904397	Soil	53/19	Left	Boundary	1074.24[2]
E08242A	WJX	902275	Dust-Vacuum	10/12	Kitchen	Inside Entryway	9.24

aThe symbol "<" means that the sample had lead below the instrument detection limit (IDL), and based on the IDL the level of lead present is less than the value given after the "<" symbol.

Table 8-1. Continued

Field Blank Outliers

Instrument Batch	Sample Preparation Batch	Lab ID	Sample Medium	Study ID/ Sample ID	Location	Component	Amount of Lead[a] (ug/sample)
E04292A	WIO	902825	Dust-Vacuum	18/06	Kitchen	Field Blank	< 0.344
E05272A	WIV	904161	Soil	70/22	Front	Field Blank	35.638
E06112A	WIW	904333	Soil	50/22	Right	Field Blank	271.625[3]
E06152A	WJU	903654	Dust-Vacuum	07/06	Living Room	Field Blank	2.682
E08032A	WKG	905133	Dust-Wipe	94/28	Kitchen	Field Blank	35.445
E08242A	WIT	904183	Soil	99/22	Front	Field Blank	< 1.197

Trip Blank Outliers

Instrument Batch	Lab ID	Sample Medium	Study ID/ Sample ID	Location	Component	Sample Weight (g)
TRIPBLNK	902217	Dust-Vacuum	19/23	Bedroom #1	Trip Blank	-0.0052
TRIPBLNK	902516	Dust-Vacuum	90/23	In Van	Trip Blank	0.0051
TRIPBLNK	902964	Dust-Vacuum	40/23	Living Room	Trip Blank	0.0002
TRIPBLNK	903144	Dust-Vacuum	07/23	Living Room	Trip Blank	0.0007
TRIPBLNK	903146	Dust-Vacuum	65/23	Living Room	Trip Blank	0.0009
TRIPBLNK	903722	Dust-Vacuum	55/23	Living Room	Trip Blank	0.0015

aThe symbol "<" means that the sample had lead below the instrument detection limit (IDL), and based on the IDL the level of lead present is less than the value given after the "<" symbol.

[1]Value subsequently corrected to 271.625 µg/g - no longer an outlier.
[2]Value subsequently corrected to 1072.76 µg/g - still an outlier.
[3]Value subsequently corrected to <5.49 - no longer an outlier.

Table 8-2. CAP Study Outliers - Laboratory QC Samples

Spike Recovery Outliers

Instrument Batch	Sample Preparation Batch	Sample ID	Run Number	Sample Type Flag	Spike % Recovery
E04272A	WIL	903695	102	2	128.5
E04272A	WIL	903701	104	3	134.0
E05042A	WIR	903551	31	2	104.1
E05042A	WIR	903555	33	3	104.0
E05072B	WJB	903604	34	2	101.5
E05072B	WJB	903597	42	3	101.5
E05072B	WJD	903584	116	2	97.8
E05072B	WJD	903753	118	3	97.9
E05122B	WJE	903454	110	2	101.2
E05122B	WJE	903484	112	3	101.2
E05192A	WIP	904266SPD	33	3	130.9
E05272A	WJO	903360	115	2	98.5
E05272A	WJO	903628	116	3	98.4
E06042A	WJP	903320	29	2	100.6
E06042A	WJP	903321	30	3	100.3
E07142A	WKF	905240	45	2	99.2
E07212A	WJC	903546	234	3	113.7
E07272A	WKJ	903303	148	2	108.5
E07272A	WKJ	903079	149	3	109.0

Method Blank Outliers

Instrument Batch	Sample Preparation Batch	Sample ID	Run Number	Sample Type Flag	Amount of Lead[a] (µg/sample)
E07272A	WIZ	MB1	38	4	<4.0202
E07272A	WIZ	MB2	39	4	<4.0202
E07272A	WKJ	MB1	116	4	4.0380
E07272A	WKJ	MB2	142	4	20.6810

[a] The symbol "<" means that the sample had lead below the instrument detection limit (IDL), and based on the IDL the level of lead present is less than the value given after the "<" symbol.

Table 8-2. Continued

Reference Material Recovery Outliers

Instrument Batch	Sample Preparation Batch	Sample ID	Run Number	Sample Type Flag	Reference Material % Recovery
E06292A	WIX	904326	181	5	114.8
E07302A	WKJ	902699	156	5	34.4
E08212A	WKJ	902699	28	5	22.9
E08212A	WIZ	902731	29	5	27.0

Continuing Calibration Blank Outliers

Instrument Batch	Sample Preparation Batch	Sample ID	Run Number	Sample Type Flag	Amount of Lead (µg/ml)
E05152A	WIK	CCB	44	9	0.0130
E05152A	WIK	CCB	93	9	0.0111
E08182A	REF	CCB	55	9	0.0004

Often, the minimum and/or maximum data values in a group were flagged as outliers by the test described above. If the minimum and maximum values in a group were not flagged, they were nevertheless included in Tables 8-1 and 8-2 as being potential outliers. Of the 838 lead loading values reported, nine (1%) were listed as potential outliers. This includes 7 out of 770 vacuum samples and 2 out of 68 wipe samples. Of the 1124 lead concentrations reported, 24 (2%) were listed as potential outliers. This includes 20 out of 770 vacuum samples and 4 out of 354 soil samples. Of the 139 field blanks, six (4%) were listed as potential outliers, and of the 53 trip blanks, six (11%) were listed as potential outliers.

8.4 RESOLUTION OF OUTLIER QUESTIONS

Tables 8-1 and 8-2 were sent to the laboratory for review. This review resulted in corrections to three of the identified field sample outliers (as indicated in footnotes to Table 8-1) and two other values which had not been identified as outliers. Two of the three outliers had similar laboratory sample ID numbers which were inadvertently switched during instrument analysis. The third outlier and the two other values were originally reported with incorrect sample weights due to re-preparation of a batch of samples. No errors were found in the reporting of the laboratory QC sample data.

8.5 DATA CERTIFICATION

In addition to the investigation of statistical outliers, an audit of the data management system was performed. In this audit 53 (out of 1413) field samples and 28 (out of 1295) laboratory QC samples were randomly selected, and all of the information in the CAPS data base for these samples was exhaustively checked against the appropriate original data sources, that is, the orginal

field data collection forms, laboratory analytical data reports, and HUD Demonstration data sets. The random selection of audit samples was constrained so that all 52 housing units, all 28 laboratory analytical batches, and all different sample types were proportionately represented.

The data management audit found no problems with any of the key data used in the statistical analysis to draw conclusions for the CAP Study. Minor problems with other information in the CAPS data base were discovered by the data management audit, such as spelling and grammatical problems in comments on field forms. These minor problems did not affect data collected from the field, nor the statistical analysis.

The laboratory which was responsible for the chemical analysis of the data used in this study also performed a quality assurance audit of the data produced by the laboratory. A total of 17.6 percent of the total samples in each batch were selected for audit. Field samples, lab QC samples, and instrument calibration samples were included. In all, 692 samples were audited, and 28 samples were found to have errors. This provides an estimated error rate of 4.05 percent, with a 95 percent confidence interval of 2.58 to 5.51 percent. The distribution of errors was as follows:

- 8 mistakes in sample identification numbers,
- 6 mistakes in dilution factors,
- 7 mistakes in weights,
- 2 mistakes in instrumental response,
- 2 mistakes in entering information, and
- 3 calculation mistakes.

The error rate found suggests an that 129 errors may be present in the remaining 3197 samples not audited. However, 100 percent verifications were later performed for sample

identification numbers and instrumental responses, correcting additional errors of these types. Although 100 percent verification was not found to perfectly correct all errors, the number of oversights is expected to be small.

In light of the 100 percent checks performed on the sample identification numbers and instrumental responses, the revised estimated error rate in the 3197 unaudited samples is 2.75 percent. This implies a total of 88 samples with errors. The upper confidence bound on this estimate is 127 samples. Restricting to field samples results in an estimate of 32 field samples with errors and an upper confidence bound of 46 errors in the field samples.